
Spis treści

Wprowadzenie.....	xi
Podziękowania	xix
1. Uzyskiwanie wartości biznesowych za pomocą AI	1
Jak AI zmienia krajobraz BI	1
Typowe przypadki użycia AI dla rozwiązań BI	7
Automatyzacja i łatwość użycia	8
Lepsze prognozowanie i przewidywania	10
Posługiwanie się danymi niestrukturyzowanymi	12
Uzyskanie intuicyjnego zrozumienia AI i uczenia maszynowego.....	13
Mapowanie przypadków użycia AI na efekty biznesowe	16
Podsumowanie	20
2. Od BI do DI: ocena wykonalności projektów AI	21
Zacznijmy od danych.	21
Ocenianie gotowości danych za pomocą metodologii 4V	23
Łączenie tych czterech cech w celu oceny gotowości danych	25
Wybór pomiędzy utworzeniem a zakupem usług AI	27
AI jako usługa.	28
Platforma jako usługa	29
Infrastruktura jako usługa	30
Pełna własność	31
Podstawowe architektury systemów AI.....	32
Warstwa użytkownika	33
Warstwa danych.....	34
Warstwa analizy	34
Uwarunkowania etyczne.....	37
Tworzenie priorytetyzowanej mapy drogowej przypadku użycia.....	40
Mieszanie czempionów i szybkich zwycięstw	42
Identyfikowanie wspólnych źródeł danych	42
Budowanie porywającej wizji.....	42
Podsumowanie	43

3. Podstawy uczenia maszynowego	45
Proces nadzorowanego uczenia maszynowego	45
Krok 1: Zbieranie danych historycznych	46
Krok 2: Identyfikowanie cech i etykiet	46
Krok 3: Podział danych na zbiory treningowe i testowe.	47
Krok 4: Użycie algorytmów do znalezienia najlepszego modelu	47
Krok 5: Ewaluacja finalnego modelu	48
Krok 6: Wdrożenie.	48
Krok 7: Konserwacja	48
Popularne algorytmy uczenia maszynowego	49
Regresja liniowa	49
Drzewa decyzyjne	50
Metody uczenia zespołowego.	52
Głębokie uczenie	54
Przetwarzanie języka naturalnego	54
Widzenie komputerowe	55
Uczenie przez wzmacnianie	55
Ewaluacja modelu uczenia maszynowego	55
Ocenianie modeli regresji.	56
Ocenianie modeli klasyfikacyjnych.	58
Ocenianie modeli klasyfikacji wielokrotnej.	61
Typowe pułapki uczenia maszynowego	62
Pułapka 1: Używanie uczenia maszynowego, gdy go nie potrzebujemy	62
Pułapka 2: Nadmierna zachłanność	62
Pułapka 3: Budowanie nadmiernie złożonych modeli	62
Pułapka 4: Nie przerywanie, gdy mamy już dość danych	63
Pułapka 5: Wpadanie w klątwę wymiarowości	64
Pułapka 6: Ignorowanie wartości odstających.	66
Pułapka 7: Przyjmowanie infrastruktury chmurowej jako pewnika	67
Podsumowanie	67
4. Prototypowanie	69
Czym jest prototyp i dlaczego jest ważny?	69
Prototypowanie w analityce biznesowej	71
Przybornik prototypowania używany w tej książce	73
Praca z Microsoft Azure	74
Rejestrowanie się w Microsoft Azure	74
Tworzenie obszaru roboczego Azure Machine Learning Studio	77
Tworzenie zasobu obliczeniowego Azure.	82
Tworzenie zasobu magazynowego.	85
Praca z Microsoft Power BI	88
Podsumowanie	89

5. Analityka opisowa wspomagana przez AI	91
Przypadek użycia: odpytywanie danych w języku naturalnym	91
Sformułowanie problemu	92
Przegląd rozwiązania.	92
Realizacja w Power BI	94
Przypadek użycia: Podsumowywanie danych w języku naturalnym	106
Sformułowanie problemu	106
Przegląd rozwiązania.	106
Realizacja w Power BI	107
Podsumowanie	112
6. Analityka diagnostyczna wspomagana przez AI	113
Przypadek użycia: automatyzowanie spostrzeżeń	113
Sformułowanie problemu	113
Przegląd rozwiązania.	114
Realizacja w Power BI	115
Podsumowanie	131
7. Analityka predykcyjna wspomagana przez AI	133
Wymagania wstępne	134
Zbiór danych.	135
Przypadek użycia: Automatyzowanie zadań klasyfikacyjnych	136
Sformułowanie problemu	136
Przegląd rozwiązania.	140
Trenowanie modelu w Microsoft Azure	141
Czym jest zadanie AutoML?	146
Ocena wyników AutoML	152
Wdrażanie modelu w Microsoft Azure	158
Uzyskiwanie przewidywań modelu za pomocą Pythona lub R.	163
Wnioskowanie z modelu w Power BI	167
Budowanie pulpitu wspomagane przez AI w Power BI	171
Przypadek użycia: Usprawnianie przewidywania KPI	174
Sformułowanie problemu	174
Przegląd rozwiązania.	177
Trenowanie modelu w Microsoft Azure	178
Wdrażanie modelu w Microsoft Azure	187
Uzyskiwanie przewidywań modelu za pomocą Pythona lub R.	188
Wnioskowanie z modelu w Power BI	191
Budowanie pulpitu wspomagane przez AI w Power BI	194
Przypadek użycia: Automatyzowanie wykrywania anomalii.	199
Sformułowanie problemu	200
Przegląd rozwiązania.	201
Włączanie usługi AI w Microsoft Azure	203

Uzyskiwanie przewidywań modelu za pomocą Pythona lub R.....	207
Wnioskowanie z modelu w Power BI	210
Budowanie pulpitu wspomaganego przez AI w Power BI.....	212
Podsumowanie	223
8. Analizy zaleceń wspomagane przez AI.....	225
Przypadek użycia: Rekomendowanie najlepszego kolejnego działania	225
Sformułowanie problemu	225
Przegląd rozwiązania.....	226
Konfigurowanie usługi AI	228
Jak działa uczenie ze wzmacnianiem w usłudze Personalizer.....	231
Konfigurowanie Azure Notebooks	233
Symulowanie interakcji z użytkownikami	235
Wykonywanie symulacji za pomocą Pythona	237
Ewaluacja wydajności modelu w portalu Azure	239
Wnioskowanie z modelu w Power BI	242
Budowanie pulpitu wspomaganego przez AI w Power BI.....	245
Podsumowanie	248
9. Wykorzystywanie nieustrukturyzowanych danych za pomocą AI	249
Przypadek użycia: Wyciąganie wniosków na podstawie danych tekstowych	249
Sformułowanie problemu	250
Przegląd rozwiązania.....	250
Konfigurowanie usługi AI	251
Konfigurowanie potoku danych	255
Wnioskowanie z modelu przy użyciu Power BI.....	264
Budowanie wspomaganego przez AI pulpitu nawigacyjnego w Power BI ...	266
Przypadek użycia: Przetwarzanie dokumentów za pomocą AI.....	268
Sformułowanie problemu	268
Przegląd rozwiązania.....	270
Konfigurowanie usługi AI	271
Konfigurowanie potoku danych	273
Wnioskowanie z modelu przy użyciu Power BI.....	281
Budowanie wspomaganego przez AI pulpitu nawigacyjnego w Power BI ...	284
Przypadek użycia: Zliczanie obiektów na zdjęciach	285
Sformułowanie problemu	285
Przegląd rozwiązania.....	286
Konfigurowanie usługi AI	287
Konfigurowanie potoku danych	287
Wnioskowanie z modelu przy użyciu Power BI.....	290
Budowanie wspomaganego przez AI pulpitu nawigacyjnego w Power BI ...	292
Podsumowanie	294

10. Zebranie wszystkiego razem: Budowanie pulpitu wspomaganego przez AI	295
Sformułowanie problemu	295
Przegląd rozwiązania	296
Przygotowywanie zbiorów danych	298
Alokowanie zasobu obliczeniowego	300
Budowanie przepływu pracy uczenia maszynowego	301
Dodawanie danych nastrojów do przepływu pracy	308
Wdrażanie przepływu pracy na potrzeby wnioskowania	316
Budowanie wspomaganego przez AI pulpitu w Power BI	322
Wykrywanie anomalii	324
Analiza predykcyjna	325
Analiza opisowa wspomaganą przez AI	326
Dane niestrukturyzowane	328
Podsumowanie	329
11. Kolejne kroki: od prototypu do produkcji	331
Odkrywanie kontra dostarczanie	331
Kryteria sukcesu dla dostarczania produktu AI	333
Ludzie	334
Procesy	336
Dane	337
Technologia	338
MLOps	340
Zaczynamy od dostarczania kompletnych inkrementacji	342
Podsumowanie	342
Indeks	345
O autorze	356

Wprowadzenie

Popkultura miała wielki wpływ na to, jak postrzegamy sztuczną inteligencję. Jeśli takie filmy SF, jak *Matrix* lub *Terminator* czegoś nas nauczyły, to tego, że AI przybywa, aby nas kontrolować (lub zniszczyć) i że ma pewne ludzkie cechy. Terminator Arnolda Schwarzeneggera pokazał, jak mogłaby wyglądać sztuczna inteligencja zamieniona w człowieka. Z *Matrixa* wynieśliśmy przekonanie, że możemy nawet nie rozpoznawać tego, że nasze życie jest w pełni kontrolowane przez przepotężną sztuczną inteligencję.

Choć w moim przekonaniu oba te wnioski są co najmniej problematyczne, zawierają nieco prawdy, jeśli chodzi o *business intelligence* (BI). AI *nadchodzi* po nas i *może* mieć ludzkie cechy. Zanim jednak przejdziemy do szczegółów, powinniśmy wyjaśnić sobie kilka kluczowych pojęć. Czym dokładnie jest BI? BI jest trudne do zdefiniowania, gdyż odpowiedź zależy od tego, kogo zapytamy, kiedy i w jakim kontekście.

Definicja BI, przynajmniej w kontekście tej książki, jest jednak całkiem prosta. Definiuję BI z systemowego punktu widzenia:

BI to system lub oprogramowanie pozwalające użytkownikom biznesowym i analitykom badać i oglądać dane z wielu źródeł w celu podejmowania decyzji z większą wiedzą.

Ilekcraz zatem odwołuję się do BI, termin ten jest blisko powiązany z systemami BI, takimi jak Microsoft Power BI, Tableau czy Looker, a także leżącymi w tle infrastrukturami, na których działają te systemy.

Zajmijmy się teraz koncepcją AI (*artificial intelligence*), który to termin będziemy eksplorować znacznie bardziej szczegółowo w tej książce. Po pierwsze, AI napędza dziś wszelkiego rodzaju procesy biznesowe. Ponad 50% wszystkich (dużych) firm na świecie wkroczyło na drogę AI z przynajmniej jakimiś niewielkimi przypadkami użycia lub wczesnymi projektami prototypowymi. Najpopularniejsze zastosowania AI obejmują:

- Mechanizmy wyszukiwania
- Prognozowanie rotacji klientów
- Przewidywanie popytu

- Przetwarzanie dokumentów
- Konserwowanie predykcyjne
- Grupowanie (segmentowanie) i personalizacja klientów
- Kontrola jakości i monitorowanie procesów w produkcji przemysłowej

Jeśli spojrzymy na rozpiętość zastosowań AI, naiwnością byłoby sądzić, że wpływy inicjatyw związanych ze sztuczną inteligencją ominą dziedzinę BI. Jest to nieuniknione i jest to tylko kwestia czasu, gdy pewna część tradycyjnych zadań BI, takich jak planowanie i przewidywanie, będzie zdecydowanie lepiej wykonywana przez AI. Jeśli ktoś stara się pokonać AI w wyszukiwaniu wzorców w danych, życzę szczęścia! Prawie na pewno przegra.

Po drugie, gdy myślimy o sztucznej inteligencji, często pojawia się wizja technologii eliminującej nasze miejsce pracy. Jednak czy takie obawy są realistyczne? Zastanówmy się, kiedy możemy oczekiwać, że zautomatyzowany system AI albo robot będzie wystarczająco dobry, aby przejąć 100% naszych codziennych obowiązków? Jeszcze pięć lat? Dziesięć? Nigdy? Z drugiej strony, kiedy kolega siedzący przy biurku obok będzie w stanie użyć AI, aby pokonać nas w ważnych zadaniach? Za rok? Za miesiąc? Czy taka osoba jest już obecna w firmie? Ważne jest, aby zrozumieć, że sama AI nie zastąpi nas w pracy. Prędzej będą to ludzie, którzy wiedzą, jak zaprząć AI do pracy i zmienić status quo. I tu na scenę wkraczają danetycy, czyli specjaliści od danych.

Krótko po ukuciu terminu *data science* w roku 2008 pojawiła się fala ambitnych danetyków o bardzo różnym poziomie wykształcenia, poczynając od kursów online po specjalistyczne studia podyplomowe. Byli zatrudniani przez czołowe korporacje światowe w nadziei wygenerowania znaczącego wpływu na biznes za pomocą danych. Danetyka została nawet nazwana „najbardziej seksownym zawodem stulecia” przez *Harvard Business Review*¹.

Nieoczekiwanie „tradycyjni” profesjonalści danych, tacy jak analitycy biznesowi i profesjonalści BI, znaleźli się na uboczu, podczas gdy nowi specjaliści zostali skierowani do pracy nad ekscytującymi zastosowaniami AI. Doprowadziło to do rozdziału pomiędzy obiema stronami.

Gdy sam zostałem pierwszym danetykiem w korporacji, spotkałem się z wielkimi oczekiwaniami ze strony zarządu, a jednocześnie podejrzliwymi spojrzeniami z bardziej „ustabilizowanych” działów zajmujących się danymi. I muszę przyznać, że początek to była dzika jazda. Gdy pierwszy raz zajrzałem do hurtowni danych firmy, zrozumiałem, że trudne może być generowanie nawet statycznych raportów z czegoś, co ja nazywałem „małymi danymi”. Z drugiej strony, moi koledzy od BI byli bardzo sceptyczni, jeśli chodzi o moją pracę: jak mógłbym w ogóle uzyskać z surowych danych choć jedną wiarygodną metrykę bez wsparcia całej armady specjalistów? I jak można w ogóle pracować bez ustalonego zakresu projektu i jasno określonych rezultatów?

¹ Thomas H. Davenport, DJ Patil, „Data Scientist: The Sexiest Job of the 21st Century”, *Harvard Business Review*, październik 2012, <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>.

Potrzeba było sporo czasu, wysiłku i dobrej woli z obu stron, abyśmy zaczęli doceniać wzajemnie swoje cele, wyzwania i umiejętności. Ostatecznie jednak zaskoczyło. Zespół BI pomógł mi lepiej zrozumieć, jak można konsekwentnie i spójnie w czasie obliczać metryki z uwzględnianiem trwałych zmian w systemach IT i procesach biznesowych. Ja natomiast mogłem im zasugerować techniki, które pomogły uzyskiwać lepsze przewidywania, szybciej wyciągać wnioski i w ogólności zapewniać lepsze działanie rozwiązań BI dzięki automatyzacji.

Czas zburzyć mur pomiędzy wrogimi „plemionami danych” w różnych organizacjach i działać wspólnie, aby osiągnąć doniosłe rezultaty. Napisałem tę książkę, aby pomóc tym po stronie BI zrozumieć, jak mogą skorzystać w wyników pracy danetyków (czyli modeli i algorytmów) i jak używać uczenia maszynowego (*machine learning* – ML) w celu poprawienia sprawności swoich rozwiązań BI. Ponadto zamierzam wyjaśnić odpowiednie metodologie identyfikowania dobrych zastosowań AI i tego, jak je implementować w duchu Agile, dzięki czemu Czytelnicy będą mogli od razu zacząć i przenieść swoje możliwości BI na następny poziom.

Aby to osiągnąć, nie zamierzam skupiać się nadmiernie na takich szczegółach, jak trenowanie lub wdrażanie modelu ML – jest to coś, co można z powodzeniem pozostawić danetykom. Zamiast tego zajmiemy się tym, jak skutecznie *używać* tych modeli do naszych celów i budować prototypy lub szybkie dowody koncepcji. Dzięki takim technologiom, jak AI jako usługa (AIaaS) i zautomatyzowanemu uczeniu maszynowemu (AutoML) będziemy w stanie omówić *wiele* usług AI bez nadmiernego zajmowania się ich wewnętrznymi szczegółami.

Książka ta nie zmieni Czytelników w danetyków, ale pozwoli lepiej zrozumieć ich rolę i to, jak danetyka może pomóc w ich pracy. Będziemy opierać się na dotychczasowej wiedzy i uczyć się, jak efektywnie implementować usługi AI w naszych własnych przepływach pracy i pulpitych. Po ukończeniu lektury tej książki powinniśmy być w stanie jeszcze wydajniej współpracować z kolegami od danetyki, a nawet samodzielnie wypełniać pewne zadania typowe dla tej dziedziny.

Kto powinien przeczytać tę książkę

Najwięcej korzyści z tej książki uzyskają osoby, które już są doświadczonymi profesjonalistami danych albo skupionymi na danych biznesmenami. Zakładam, że czytelnik dobrze zna dane swojej organizacji (albo przynajmniej te części danych, przy których pracuje) i jest świadomy pułapek i ograniczeń, jakie zawierają. W swojej roli jako profesjonalista BI, analityk biznesowy, analityk danych albo programista skupiony na danych rozumie wartość danych organizacji w konkretnym przypadku użycia. Tym, co nadal chce ustalić, jest to, czy i jak AI może usprawnić procesy lub zapewnić lepsze wnioski, a tym samym wnieść dodatkową wartość do naszego biznesu. Czytelnik chce zdobyć praktyczne doświadczenie, samodzielnie budując kompletny przypadek użycia AI albo przynajmniej zrozumieć go od początku do końca. Innymi słowy, nie jest jeszcze ekspertem AI i raczej nie pragnie nim być, ale chce nauczyć się więcej z tej dziedziny.

Aby uzyskać najwięcej korzyści z tej książki, nie należy obawiać się samodzielnego programowania (niewielkiego). Choć w większości przykładów będziemy trzymać się narzędzi nie wymagających kodowania, w pewnych miejscach po prostu łatwiejsze jest uruchomienie małego skryptu w Pythonie lub R, aby zagwarantować gładkie działanie. Przykładowo będziemy pobierać dane z API typu HTTP REST, aby wbudować prognozy AI do naszych pulpitów. Inna sytuacja to pewne podstawowe przetwarzanie danych za pomocą biblioteki *pandas* Pythona albo *Tidyverse* w R. Jeśli ktoś jest otwarty i chce się czegoś nauczyć z tych ćwiczeń, książka zapewni wszystko, co potrzebne, aby samodzielnie utworzyć pierwszy prototyp rozwiązania AI bez prowadzenia za rękę, a wszystko to dzięki gotowym do użycia szablonom kodu.

Choć nie zakładam zbyt dużego przygotowania technicznego, oczekuję, że czytelnicy dobrze znają podstawowe przepływy pracy w analizie danych. Przykładowo spodziewam się umiejętności analizowania danych względem wielu wymiarów, tworzenia wykresów liniowych i słupkowych czy zrozumienia tego, jak działają pliki CSV. Takie terminy, jak *statystyka opisowa* czy *regresja liniowa*, nie powinny powodować zbyt dużych obaw.

Wszystkie przykłady zostały zaprezentowane przy użyciu Microsoft Power BI. Jeśli ktoś już zna Power BI, będzie to duża przewaga. Jeśli nie, w książce przedstawiam główne koncepcje, ale zapewne warto będzie przeczytać przynajmniej jedno z dodatkowych źródeł przedstawionych w kolejnym podrozdziale.

Microsoft Power BI i Azure

Choć wszystkie koncepcje i metodologie przedstawione w książce zostały napisane tak, aby były uniwersalne, ćwiczenia są wyjaśniane w ramach pojedynczego stosu technicznego. Dla celów tej książki wykorzystuję oprogramowanie firmy Microsoft, a konkretnie Power BI Desktop jako frontend oraz Microsoft Azure jako chmurowe usługi AI. Jeśli ktoś preferuje inne narzędzia, powinien być w stanie łatwo zmienić kontekst, jako że większość interfejsów i przepływów pracy jest całkiem podobnych w innych narzędziach BI, takich jak Tableau i Qlik, a w przypadku usług chmurowych AI te oferowane przez Amazon Web Services lub Google Cloud Platform również są zbliżone.

Aby podążać za przykładami pokazanymi w książce, potrzebny będzie dostęp do następującego oprogramowania:

Microsoft Power BI Desktop

Wersja bezpłatna będzie wystarczająca dla realizacji ćwiczeń. Oprogramowanie Microsoft Power BI można pobrać z jego witryny (<http://www.powerbi.com/>). Jeśli ktoś nigdy nie posługiwał się wcześniej programem Power BI, warto będzie zapoznać się z przewodnikiem „Introduction to Power BI” (<https://oreil.ly/KseuL>). Osoby chcące lepiej poznać to narzędzie mogą sięgnąć po książkę *Poznaj Microsoft Power BI*, której autorem jest Jeremy Arnold (wyd. polskie APN Promise 2023, ISBN 978-83-7541-506-3).

Microsoft Azure Cloud Services

W tej książce pokazuję wykorzystanie usług AI dostępnych na platformie chmurowej Microsoft Azure. Czytelnik powinien być w stanie zmieścić się (z zapasem) w ramach bezpłatnej subskrypcji próbnej. Zależnie od tego, jak często będą uruchamiane ćwiczenia i wielkości własnych zbiorów danych (o ile ktoś zechce przetestować rozwiązania na własnych danych) może się jednak okazać, że granice poziomu bezpłatnego zostaną przekroczone. Warto sprawdzać kalkulator cenowy Microsoft Azure, jeśli pojawią się wątpliwości dotyczące dowolnych powiązanych kosztów.

Cele nauki

Po ukończeniu książki Czytelnik powinien rozumieć następujące zagadnienia:

- Jak AI może generować efekty biznesowe w środowiskach BI
- Jakie są najważniejsze przypadki użycia AI w analityce biznesowej
- Jak rozpocząć korzystanie z AI poprzez szybkie prototypowanie
- Jakie narzędzia prototypowania są dostępne w kontekście sztucznej inteligencji
- Jak budować rozwiązania wspomagane przez AI w kontekście BI
- Jak zbudować kompletny prototyp w celu zweryfikowania zwrotu z inwestycji w sztuczną inteligencję
- Jak przejść od prototypu do rozwiązania produkcyjnego

Powinien również opanować następujące umiejętności:

- Używanie AutoML do zautomatyzowanych klasyfikacji i lepszego prognozowania
- Implementowanie usług zaleceń jako wsparcia dla podejmowania decyzji
- Wyciąganie wniosków z danych tekstowych w wielkiej skali za pomocą usług przetwarzania języka naturalnego
- Wydobywanie informacji z dokumentów i obrazów za pomocą usług widzenia komputerowego
- Budowanie interaktywnych pulpitów dla prototypów wspomaganych przez AI
- Implementowanie kompletnych studiów przypadku dla budowania pulpitów analitycznych wspomaganych przez AI

Poruszanie się po tej książce

Po przedstawieniu tła i celów, które książka powinna pomóc osiągnąć, przyjrzyjmy się jej strukturze.

Książka podzielona jest na dwie zasadnicze części. Rozdziały 1 do 4 przedstawiają podstawy teoretyczne i przygotowanie środowiska pracy do ćwiczeń. Rozdziały 5 do 11

przedstawiają praktyczne przykłady, które demonstrują, jak budować lepsze pulpity nawigacyjne i prognozy lub odblokowywać niestrukturyzowane dane za pomocą AI.

Rozdział 1, „Uzyskiwanie wartości biznesowych za pomocą AI”, omawia zagadnienie wpływu, jaki sztuczna inteligencja może mieć na analitykę biznesową i jak typowo można wykorzystywać AI w kontekście BI. Przedstawiam w nim również ramy pozwalające priorytetyzować przypadki użycia AI/ML zgodnie z ich wpływem na biznes.

Rozdział 2, „Od BI do DI: ocena wykonalności projektów AI”, skupia się na technicznej wykonalności projektów wykorzystujących sztuczną inteligencję. Poznamy w nim typowe wzorce rozwiązań AI i uzyskamy ogólne pojęcie, jak zbudowane są usługi AI. Na koniec będziemy w stanie oszacować wykonalność projektu AI, a dzięki temu utworzyć priorytetyzowaną mapę drogową dla naszych przypadków użycia.

Rozdział 3, „Podstawy uczenia maszynowego”, przedstawia fundamentalne koncepcje ML, które nie tylko umożliwiają stosowanie technik AI, ale również pozwolą zrozumieć, jak to robią i jak to wszystko mieści się w ogólnym obrazie. Poznamy również kilka podstawowych pułapek związanych z uczeniem maszynowym.

Rozdział 4, „Prototypowanie”, jest wprowadzeniem do jednego z najpotężniejszych narzędzi zwinnego zarządzania projektami (w ogólności) i danetyki (w szczególności): prototypowania. Dowiemy się z niego, czym są prototypy, dlaczego są ważne i jak budować skuteczne prototypy systemów AI.

Rozdział 5, „Analityka opisowa wspomagana przez AI”, to początek części drugiej książki – praktycznych ćwiczeń. Dowiemy się w nim, jak implementować funkcjonalności AI, aby szybciej realizować analizy opisowe i zapewnić bardziej intuicyjne i płynne sposoby interakcji z wielkimi zbiorami danych.

Rozdział 6, „Analityka diagnostyczna wspomagana przez AI”, idzie krok dalej. Zajmiemy się tu nie tylko opisywaniem danych, ale nadawaniem im większego sensu dzięki wspomaganemu analiz diagnostycznych i automatycznemu wykrywaniu ciekawych wzorców, tak by móc więcej czasu poświęcić na interpretowanie danych.

W rozdziale 7, „Analityka predykcyjna wspomagana przez AI”, dowiemy się, jak wyjść poza wnioski z informacji historycznych i implementować wspomagane przez AI analizy predykcyjne. Rozdział ten zawiera różne przykłady, od klasyfikacji po przewidywania i wykrywanie anomalii.

Rozdział 8, „Analizy zaleceń wspomagane przez AI”, to kolejny krok, w którym nie będziemy jedynie przewidywać wyników, ale również sugerować najlepsze kolejne działania, które należy podjąć, wykorzystując system rekomendacji.

W rozdziale 9, „Wykorzystywanie niestrukturyzowanych danych za pomocą AI”, dowiemy się ostatecznie, jak wyjść poza dziedzinę danych tabelarycznych. Będziemy w nim badać, jak AI może pomóc w automatyzowaniu przetwarzania danych niestrukturyzowanych, takich jak teksty, dokumenty i pliki obrazów.

Rozdział 10, „Zebranie wszystkiego razem: Budowanie pulpitu wspomaganego przez AI”, łączy ze sobą wiedzę wyniesioną z wcześniejszych rozdziałów. Zbudujemy w nim kliencki pulpit analityczny wspomagany przez AI, oparty na indywidualnych blokach konstrukcyjnych poznanych w poprzednich rozdziałach.

Po świętowaniu sukcesu naszego pierwszego dowodu poprawności koncepcji rozwiązania BI wspomaganego przez AI, rozdział 11, „Kolejne kroki: od prototypu do produkcji”, omawia kolejne kroki konieczne do przejścia z fazy prototypu do rozwiązania produkcyjnego i podsumowuje wszystko, czego się nauczyliśmy.

Książce towarzyszy witryna internetowa, w której można znaleźć wszystkie pliki demonstracyjne i fragmenty kodu użyte w ćwiczeniach praktycznych. Warto utworzyć sobie zakładkę dla tej strony. W każdym rozdziale znajdują się bezpośrednie odsyłacze do zasobów w tej witrynie.

Gotowi? Kluczyki w dłoń i siadamy na przednim fotelu. Zajmę miejsce obok, prowadząc Czytelników poprzez świat analityki biznesowej wspomaganą sztuczną inteligencją.

Konwencje używane w książce

W tej książce wykorzystywane są następujące konwencje typograficzne:

Kursywa

Jest używana do oznaczania nowych terminów, adresów URL, nazw plików i rozszerzeń plików.

Czcionka o stałej szerokości

Jest używana w listingach programów, a także do oznaczenia elementów programu, takich jak nazwy zmiennych lub funkcji, baz danych, typów danych, zmiennych środowiskowych i słów kluczowych.

Pogrubiona czcionka o stałej szerokości

Pokazuje polecenia lub inny tekst, który użytkownik powinien wpisać.

Pochyła czcionka o stałej szerokości

Pokazuje tekst, który należy zastąpić wartością podaną przez użytkownika lub zależną od kontekstu.



Ten element oznacza wskazówkę lub sugestię.



Ten element oznacza ogólną uwagę.



Ten element jest ostrzeżeniem.

Używanie przykładów kodu

Dodatkowe materiały dla tej książki (przykłady kodu, ćwiczenia, pliki danych) można pobrać ze strony <https://www.aipoweredbi.com>.

Jeśli pojawią się pytania techniczne lub problemy przy używaniu przykładów kodu, prosimy o zgłaszanie ich emailem na adres bookquestions@oreilly.com.

Ta książka ma pomagać w pracy. Przykładowy kod można wykorzystać w swoich programach i dokumentacji. Czytelnik nie musi nas prosić o pozwolenie, o ile nie zamierza wykorzystać znacznej części kodu. Przykładowo, napisanie programu zawierającego kilka fragmentów kodu z tej książki nie wymaga pozwolenia. Sprzedaż lub dystrybucja przykładów z książek wydawnictwa O'Reilly wymaga pozwolenia. Odpowiedź na pytanie zawierające cytaty z tej książki oraz przykładowy kod nie wymaga pozwolenia. Umieszczenie znacznej ilości przykładowego kodu z tej książki w dokumentacji produktu wymaga pozwolenia.

Doceniamy, ale nie wymagamy atrybucji. Atrybucja zwykle obejmuje tytuł, autora, wydawnictwo i ISBN. Przykładowo: „*Analityka biznesowa wspomagana sztuczną inteligencją*, Tobias Zwingmann. Copyright 2022 Tobias Zwingmann, 978-83-7541-510-0”.

Jeśli ktoś uważa, że będzie używał przykładów kodu w sposób wymagający uzyskania pozwolenia, powinien się z nami skontaktować pisząc pod adres permissions@oreilly.com.

Jak się z nami skontaktować?

Wszelkie komentarze i pytania związane z tą książką proszę przesyłać do wydawcy:

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

800-998-9938 (w Stanach Zjednoczonych lub Kanadzie)

707-829-0515 (międzynarodowy lub lokalny)

707-829-0104 (faks)

Mamy dedykowaną stronę dla tej książki, na której zamieściliśmy erratę, przykłady i wszelkie dodatkowe informacje. Znajdziesz ją pod adresem <https://oreil.ly/ai-powered-bi>.

Aby skomentować lub zadać pytanie techniczne dotyczące tej książki, wyślij e-mail na adres bookquestions@oreilly.com.

Więcej informacji na temat naszych książek i kursów znajdziesz na stronie <https://oreilly.com>.

Znajdź nas na LinkedIn: <https://linkedin.com/company/oreilly-media>.

Obserwuj nas na Twitterze: <https://twitter.com/oreillymedia>.

Oglądaj nas na YouTube: <https://youtube.com/oreillymedia>.

Podziękowania

Pisanie książki to sposób, jakim mężczyzna najbardziej może się zbliżyć do aktu urodzenia dziecka.

—*Norman Mailer*

Ten cytat bardzo mnie porusza. Nie tylko dlatego, że pisanie tej książki zajęło mi niemal dokładnie dziewięć miesięcy. Przede wszystkim dlatego, że przypomina, komu muszę podziękować w pierwszej kolejności i najbardziej: mojej rodzinie.

Pisanie oznacza poświęcanie. Poświęcamy wiele czasu, energii i innych rzeczy na rzecz mnóstwa słów na papierze, zaś skutek jest mocno niepewny. Nie możemy przyjmować za pewnik, że rodzina pomoże nam w tak nedorzecznym przedsięwzięciu. Mam to szczęście, że zawsze mogę liczyć na wsparcie mojej cudownej żony Çiğdem. Jej miłość, zaufanie i optymizm sprawiają, że wszystko to było możliwe i inspiruje mnie każdego dnia. Proste „dziękuję” nie może wyrazić, co czuję. Jesteś najlepsza. Kocham cię.

Każde z nas byłoby nikim bez naszych matek. Dlatego chciałbym podziękować mojej mamie Anett i teściowej Gülten za to, że zawsze stoją przy nas i nigdy nas nie zawiodły. Gdyby nie one, nie byłoby możliwe podążanie za marzeniami i jednocześnie wychowywanie trójki cudownych dzieci.

Chciałbym też podziękować całemu zespołowi redakcyjnemu w O'Reilly za to, że ta książka była możliwa. W szczególności dziękuje Michelle Smith za wsparcie i wiarę we mnie. Chcę też podziękować redaktorze merytorycznej, Ricie Fernando, za cenne uwagi, komentarze i ogólną pomoc. Bez jej udziału wielokrotnie zboczyłbym na mieliznę! Ostatni, choć nie najmniej ważny, jest Chris Faucher, który przekształcił stertę rękopisów w tak piękną książkę i tchnął w nią życie!

Wiele ścieżek może prowadzić do napisania książki. Moja rozpoczyna się od spotkania z George'm Mount. Gdybyśmy się nie poznali, gdyby nie jego wzmianka o takim pomysle, gdyby nie zachęcał mnie do pisania, ta książka nigdy by nie powstała. Jest to piękny przykład tego, że rzeczy wielkie zaczynają się od małych kroków. Jestem bardzo wdzięczny, że ten pierwszy krok zrobiłem z tobą, George!

Redagowanie książki technicznej jest niemal równie trudne, jak jej pisanie. Trójka redaktorów – George Mount, Donald Farmer i Michael Norris – zasługuje na wielkie uznanie za krytyczne czytanie treści i cenne uwagi, z wielką dbałością o szczegóły.

Książka ta obejmuje tak wiele zagadnień i koncepcji, że niemożliwe byłoby, aby ktoś pojedynczy znał je wszystkie. Wiele osób pracowało ze mną albo przekazywało swoje spostrzeżenia. Należą do nich Ram Kumar, Alexander Niltop, Franco Arda, Piotr Menclewicz, Felix Urban, Marek Drob i wielu innych. Dziękuję za to wszystko!

Na koniec chciałbym podziękować wszystkim, którzy wnoszą wkład do świata oprogramowania open source w postaci takich projektów, jak Python, R i wszystko, co przychodzi z nimi. Bez waszej pracy u podstaw po prostu nie byłoby możliwe wciśnięcie tak wielu pomysłów w jedną praktyczną książkę.

Uzyskiwanie wartości biznesowych za pomocą AI

W tym rozdziale zbadamy to, dlaczego włączenie sztucznej inteligencji (*artificial intelligence*, AI) do analiz biznesowych (*business intelligence*, BI) staje się ważniejsze, niż kiedykolwiek i jak zespoły BI mogą wykorzystywać AI. W tym celu zidentyfikujemy typowe obszary, w których AI może wspomagać zadania i procesy BI, a także przyjrzymy się leżącym w tle funkcjonalnościom uczenia maszynowego (*machine learning* – ML). Na koniec tego rozdziału przejdziemy do praktycznej platformy roboczej, która pozwoli odwzorować funkcjonalności AI/ML na różne obszary problemów BI.

Jak AI zmienia krajobraz BI

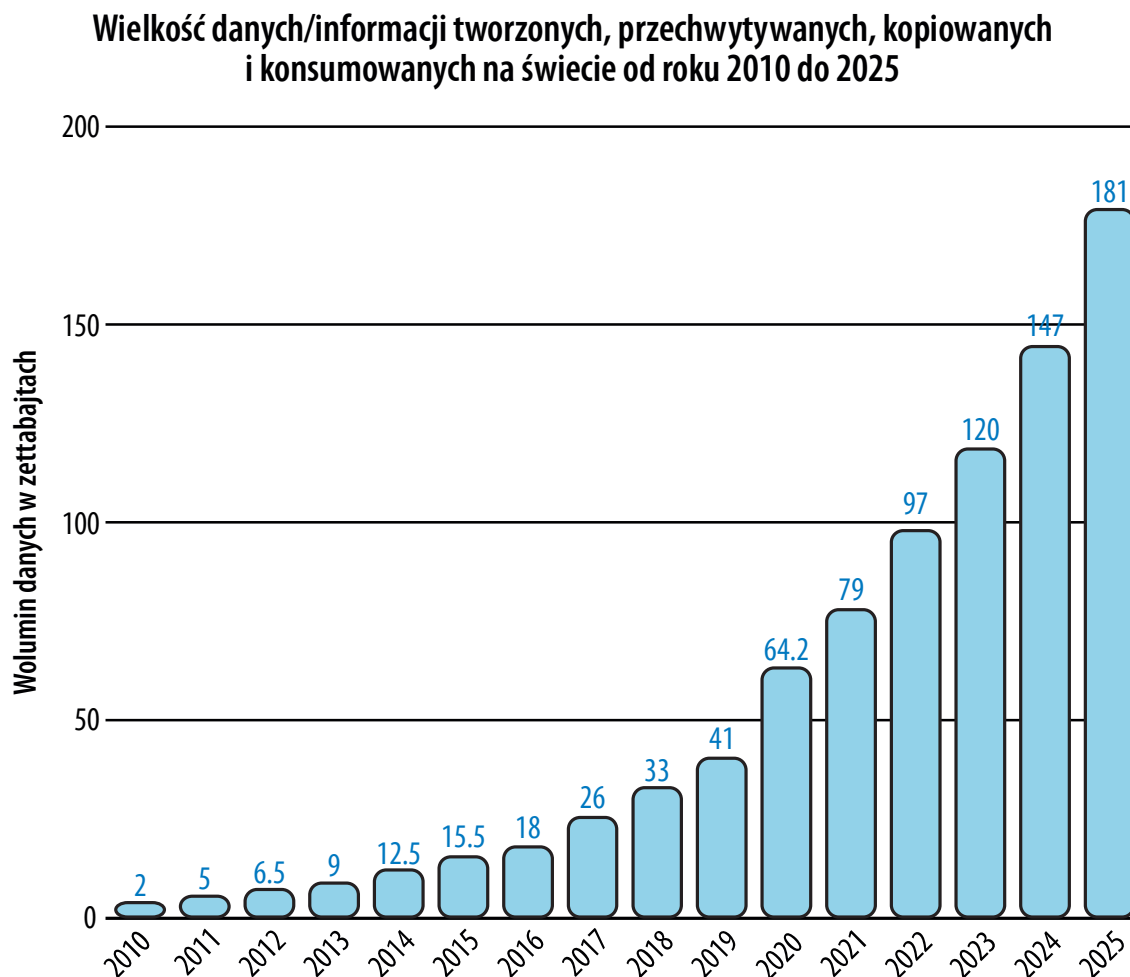
W ciągu ostatnich 30 lat BI powoli, ale konsekwentnie stawało się siłą napędową leżącą u podstaw opartej na danych kultury firmowej – a przynajmniej do momentu, gdy centrum uwagi przesunęło się w stronę danetyki, uczenia maszynowego i AI. Jak się to w ogóle stało? I co to oznacza dla struktury naszych BI?

Spoglądając wstecz na początki pierwszej ery systemów wspomagania decyzji w latach siedemdziesiątych XX wieku, ujrzymy systemy techniczne wykorzystywane przez ekspertów informatycznych do uzyskiwania wniosków na podstawie niewielkich (z dzisiejszego punktu widzenia) zbiorów danych. Systemy te ewoluowały, aż ostatecznie zaczęto je określać terminem *BI* w późnych latach 80. Analizowanie danych było czymś zupełnie nowym, zatem nawet najbardziej podstawowe spostrzeżenia budziły podziw. Nagle okazało się, że decyzje biznesowe nie są już oparte jedynie na wewnętrznym instynkcie, ale na rzeczywistych danych, które w złożonych sytuacjach pozwoliły podejmować bezpieczniejsze i odważniejsze decyzje.

Druga era rozwoju BI rozpoczęła się na przełomie wieków i dominowały w niej rozwiązania określane jako analizy samoobsługowe (*self-service*). Bogactwo nowych narzędzi i technik sprawiało, że fragmentowanie i selekcjonowanie danych przez nietechnicznych

użytkowników stało się łatwiejsze, niż kiedykolwiek. Nagle okazało się, że mogą oni bez większego wysiłku wybierać dane, tworzyć wizualizacje i wyciągać wnioski z coraz większych źródeł danych. Rozwiązania te były głównie oferowane przez wielkie firmy programistyczne, takie jak Oracle, SAP i Microsoft, ale również pobudziły rozwój niszowych firm skoncentrowanych na BI, takich jak Tableau Software. Coraz silniejsza była też integracja powszechnie używanych arkuszy kalkulacyjnych z ogólnym ekosystemem analizowania danych – na przykład pozwalając użytkownikom biznesowym (nie informatykom) na dostęp do kostek OLAP poprzez tabele przestawne w Microsoft Excel.

Większość większych firm obecnie nadal tkwi w tej drugiej fazie rozwoju BI. Dlaczego? Po pierwsze, w ostatnich latach wiele wysiłków technologicznych koncentrowało się na technicznych aspektach zarządzania wykładniczym wzrostem rozmiarów danych, które miały być przetwarzane przez systemy BI w celu wyciągania wniosków i spostrzeżeń. Po drugie, wzrost rozmiarów danych, napędzany przede wszystkim rozwojem Internetu i usług cyfrowych (patrz rysunek 1-1), doprowadził do coraz większego deficytu osób umiających posługiwać się danymi, dysponujących umiejętnościami obsługi wielowymiarowych zbiorów danych i niezbędnych do tego narzędzi (w tym przypadku *nie* Excela).



Rysunek 1-1 Wzrost wielkości danych w ostatnich latach.

Źródło: Statista (<https://oreil.ly/aYY04>)

W porównaniu z rynkiem konsumenckim, w obszarze profesjonalnego BI wykorzystywanie AI nadal jest obsługiwane niewystarczająco. Wynika to zapewne stąd, że osoby utalentowane w kierunku AI i BI zazwyczaj trafiają do różnych zespołów w ramach organizacji, a jeśli nawet się kiedyś spotykają, trudno im jest efektywnie komunikować się ze sobą nawzajem. To zaś jest efektem tego, że obie grupy typowo „mówią różnymi językami” i mają odmienne priorytety: eksperci BI zwykle nie zajmują się trenowaniem modeli i testowaniem danych, zaś danetycy rzadko mają do czynienia z pakietami SQL Server Integration Services (SSIS) i procedurami ekstrakcji, transformacji i ładowania danych (ETL).

Potrzeba włączenia technik AI do rozwiązań BI będzie jednak wzrastać w nieunikniony sposób, co można wnioskować z następujących, zauważalnych trendów:

Potrzeba uzyskiwania szybkich odpowiedzi z posiadanych danych

Organizacje domagają się spostrzeżeń opartych na danych, aby zachować konkurencyjność i się rozwijać. Analitycy danych już są przytłoczeni rosnącą liczbą żądań, aby zbadać tę lub tamtą miarę lub przeanalizować ten lub tamten zbiór danych. Jednocześnie użytkownicy biznesowi mają coraz większe potrzeby szybkiego i łatwego uzyskiwania odpowiedzi na podstawie danych. Jeśli mogą zapytać Google lub Amazon Alexa o bieżącą cenę akcji konkretnej firmy, dlaczego nie mogliby zapytać swojego profesjonalnego systemu BI o wartości sprzedaży z wczoraj?

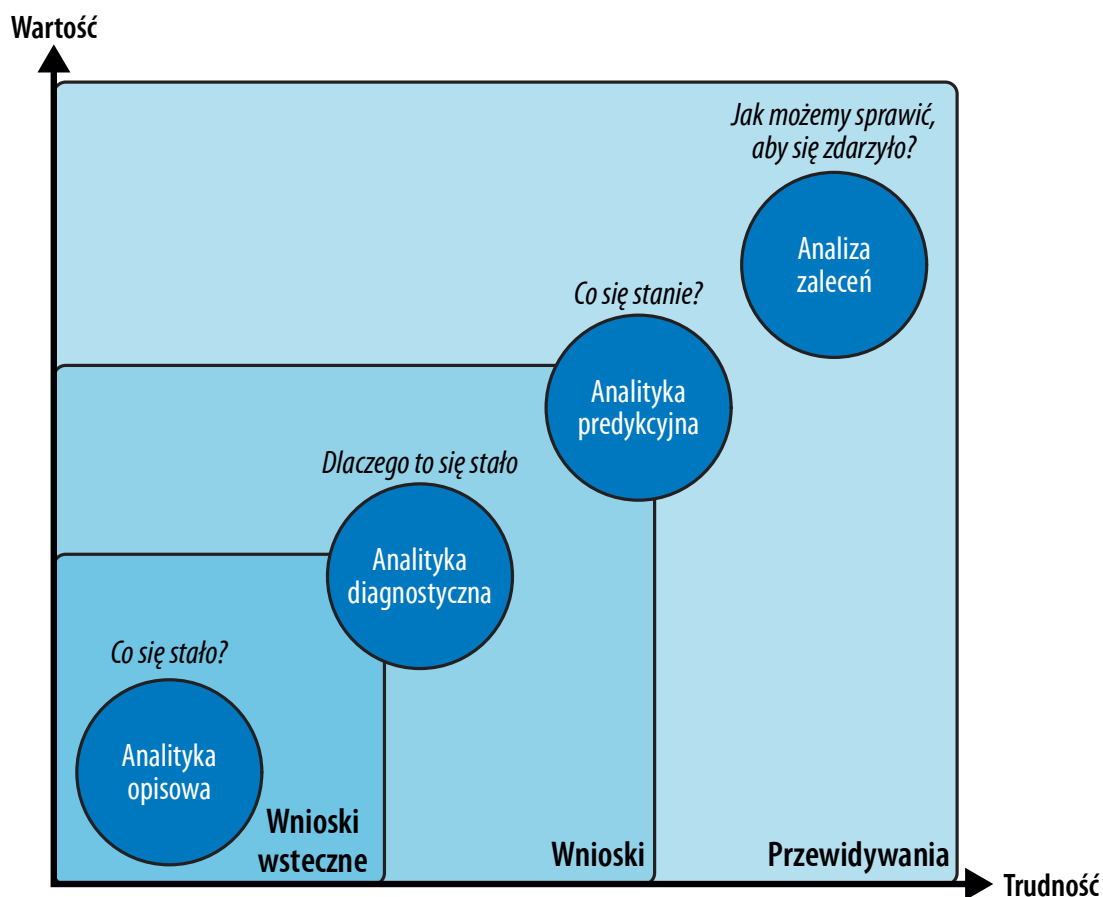
Demokratyzacja spostrzeżeń

Użytkownicy biznesowi przyzwyczaili się do uzyskiwania spostrzeżeń z danych za pomocą samoobsługowych rozwiązań BI. Jednak dzisiejsze zbiory danych są często zbyt obszerne i zbyt złożone, aby można było je przekazać do czystej samoobsługowej analizy. Rosnące objętości danych, ich zmienność i ruchliwość sprawiają, że trudne jest, jeśli nie niemożliwe, analizowanie ich przez nietechnicznych użytkowników przy użyciu znajomych narzędzi na ich własnych komputerach. Aby móc dalej demokratyzować uzyskiwanie spostrzeżeń w całej organizacji, potrzebne są systemy BI, które będą łatwe w użyciu i będą automatycznie ujawniać spostrzeżenia użytkownikom końcowym.

Dostępność usług ML

Choć wykorzystanie AI nadal rośnie w większości organizacji, jednocześnie rosną oczekiwania co do jakości przewidywań i prognoz. W jeszcze większym stopniu dotyczy to rozwiązań BI; platformy z niewielką potrzebą kodowania lub w ogóle bez takiej konieczności sprawiają, że łatwiejsze, niż kiedykolwiek staje się udostępnianie technik uczenia maszynowego osobom nie będącym specjalistami od danych i nakładają na członków zespołów BI potrzebę włączania spostrzeżeń predykcyjnych do ich raportów. Te same postępy, które obserwujemy w danetyce, są oczekiwane również w obszarze BI, prędzej lub później.

Aby lepiej zrozumieć, jak zespoły BI mogą wykorzystywać AI, przyjrzyjmy się pokrótce modelowi spostrzeżeń analitycznych opublikowanych przez firmę Gartner (<https://oreil.ly/GGxKR>, rysunek 1-2).



Rysunek 1-2 Rodzaje wniosków i metodologii analitycznych.

Źródło: Gartner (<https://oreil.ly/GGxKR>)

Kluczowa funkcjonalność każdego rozwiązania BI albo infrastruktury raportującej to dostarczanie wniosków wstecznych i prognostycznych poprzez wykorzystanie *analiz opisowych* i *diagnostycznych* danych historycznych. Te dwie metody są najważniejsze dla wszystkich dalszych procesów analitycznych, bazujących na ich wynikach.

Po pierwsze, organizacja musi zrozumieć, co wydarzyło się w przeszłości i co było motorem tych zdarzeń z punktu widzenia danych. Jest to typowo nazywane *bazowym raportowaniem* z pewnymi elementami spostrzeżeń. Trudność techniczna i złożoność jest stosunkowo niska, ale to samo dotyczy wartości wewnętrznej takich informacji. Nie należy tego źle rozumieć: rzetelne i strukturalne raportowanie nadal jest najważniejszym szkieletem analiz danych w biznesie, jako że tworzy fundament dla bardziej zaawansowanych koncepcji i generuje pytania lub problemy, które napędzają dalszą analizę. W istocie każda faza modelu spostrzeżeń obejmuje wszystkie wcześniejsze etapy. Nie można wykonać analizy predykcyjnej, dopóki nie nadamy znaczenia swoim danym historycznym.

Rozważmy następujący przykład. Firma telekomunikacyjna ma klientów, którzy subskrybują usługi w cyklu miesięcznym. W każdym miesiącu pewna liczba klientów nie odnawia swojego kontraktu usługowego i porzuca relację biznesową – zjawisko to nazywane jest *rotacją klientów* (*customer churn*).

Najbardziej podstawowym wymogiem, który musi spełnić system BI, będzie ustalenie, jak wielu klientów podlegało rotacji w przeszłości i jak rotacja zmienia się z upływem czasu. *Analityka opisowa* zapewnia nam niezbędne informacje dla ustalenia wysokości współczynnika rotacji w czasie i stwierdzenia, czy mamy do czynienia z rzeczywistym problemem. Tabela 1-1 pokazuje przykład, jak mogłoby to wyglądać.

Tabela 1-1 *Współczynnik rotacji w czasie (analityka opisowa)*

Miesiąc	Kwartał 1			Kwartał 2		
	Styczeń 22	Luty 22	Marzec 22	Kwiecień 22	Maj 22	Czerwiec 22
Współczynnik rotacji	24%	26%	22%	29%	35%	33%

Wartość wewnętrzna tej informacji jest raczej niewielka. Na tym poziomie analiza nie może naprawdę rozstrzygnąć, dlaczego wystąpiło obserwowane zjawisko ani co moglibyśmy z nim zrobić. Jednak przynajmniej sygnalizuje, czy w ogóle mamy jakiś problem: w tej tabeli możemy zauważyć, że współczynnik rotacji w kwartale 2 wydaje się znacząco wyższy, niż w pierwszym, zatem wartościowe może być bliższe przyjrzenie się temu zjawisku.

W tym momencie do gry wchodzi *analityka diagnostyczna*. Możemy teraz kopać głębiej i wzbogacić transakcyjne dane sprzedaży o dodatkowe informacje o klientach – na przykład o grupy wiekowe, co pokazuje tabela 1-2.

Tabela 1-2 *Współczynnik rotacji względem czasu i segmentów klientów*

Wiek klienta	Współczynnik rotacji w kwartale 1	Współczynnik rotacji w kwartale 2
18–29	29%	41%
30–49	28%	30%
50–64	24%	25%
65 i więcej	20%	19%

Ta analiza informuje nas, że współczynniki rotacji wydają się stabilne wśród klientów, którzy mają lat 50 i starszych. Z drugiej strony, wydaje się, że młodszy klienci chętniej dokonują rotacji i ten trend zauważalnie wzrósł w kwartale 2. Klasyczne systemy BI pozwolą na analizowanie tych danych względem wielu zmiennych, aby ustalić, co się dzieje.

W wielu przypadkach przedsiębiorstwo może znaleźć wartościowe wzorce, które jedynie dzięki analizom tego rodzaju doprowadzą do odpowiednich działań lub decyzji. To dlatego ta faza jest tak krytyczna i zawsze pozostanie bardzo ważna.

Analityka predykcyjna przenosi działania analityczne krok dalej, odpowiadając na pojedyncze pytanie: co nastąpi w przyszłości przy założeniu, że wszystkie wzorce znane z przeszłości będą się powtarzać? Tym samym analiza predykcyjna dodaje do danych kolejny poziom wartości, ale i złożoności, co widzimy w tabeli 1-3.

Tabela 1-3 Szacowanie prawdopodobieństwa rotacji klientów (analityka predykcyjna)

ID klienta	Wiek	Plan	Cena	Aktywny od (miesiące)	Prawdopod. rotacji
12345	24	A	9,95	13	87%
12346	23	B	19,95	1	95%
12347	45	B	19,95	54	30%

Złożoność się powiększa, gdy opuszczamy królestwo danych historycznych. Zamiast formułowania spostrzeżeń jako pojęć binarnych (takich jak prawda i fałsz), wprowadzamy prawdopodobieństwo wystąpienia określonych zdarzeń (w tym przypadku rotacji klientów). W tym samym czasie rośnie wartość danych, gdyż wszystko, co wiemy z przeszłości, włączamy do przypuszczeń na temat tego, jak wpływa to na przyszłe zachowania.

Dla przykładu, opierając się na prawdopodobieństwie przyszłej rotacji i historycznych danych sprzedaży możemy obliczyć prognozowane ryzyko sprzedaży dla kolejnych kwartałów, które zostanie włączone do planowania finansowego. Możemy też wybrać tych klientów, dla których prawdopodobieństwo rotacji jest wysokie, aby podjąć ukierunkowane działania w celu złagodzenia tego ryzyka.

Ale jakie działania powinniśmy podjąć? Zapraszamy do *analityki zaleceń!* Tabela 1-4 pokazuje, jak mogłoby to wyglądać w praktyce. W tym przypadku dodajemy kolejny wymiar, który nazwaliśmy *następna najlepsza oferta*; zawiera on zalecane działanie, takie jak określony rabat lub ulepszenie produktu w zależności od indywidualnego profilu klienta i jego historycznego zachowania zakupowego.

Tabela 1-4 Sugerowanie działania (analityka zaleceń)

ID klienta	Wiek	Plan	Cena	Aktywny od (miesiące)	Prawdopod. rotacji	Następna najlepsza oferta
12345	24	A	9,95	13	87%	Oferta rocznego kontraktu
12346	23	B	19,95	1	95%	Ulepszenie produktu
12347	45	B	19,95	54	30%	Brak

Gdy skupimy się na organizacjach, które mają tysiące (wiele tysięcy) klientów, staje się oczywiste, że w celu zoptymalizowania takich zadań w skali makro konieczne jest poleganie na automatyzacji w skali mikro. Ręczne podejmowanie wszystkich takich mini-decyzji i monitorowanie efektywności działań dla każdego klienta jest po prostu niewykonalne. Poziom zwrotu z inwestycji (*return on investment* – ROI) tych mini-decyzji jest po prostu zbyt niski, aby usprawiedliwić podejmowanie wysiłku.

To właśnie jest miejsce, w którym AI i BI mają doskonały punkt spotkania. Przypuśćmy, że AI może wskazać prawdopodobieństwo rotacji wraz z sugerowaną następną ofertą dla każdego klienta. Możemy teraz mieszać te informacje z klasycznymi miarami BI, takimi jak historyczne przychody od klienta albo jego lojalność, co pozwoli podjąć świadomą decyzję o tych działaniach, które będą miały największy wpływ biznesowy i największe szanse powodzenia.

Zależność pomiędzy AI i BI możemy zatem ładnie podsumować poniższym równaniem:

$$AI + BI = DI$$

gdzie *DI* oznacza *Decision Intelligence* (analityka decyzyjna)

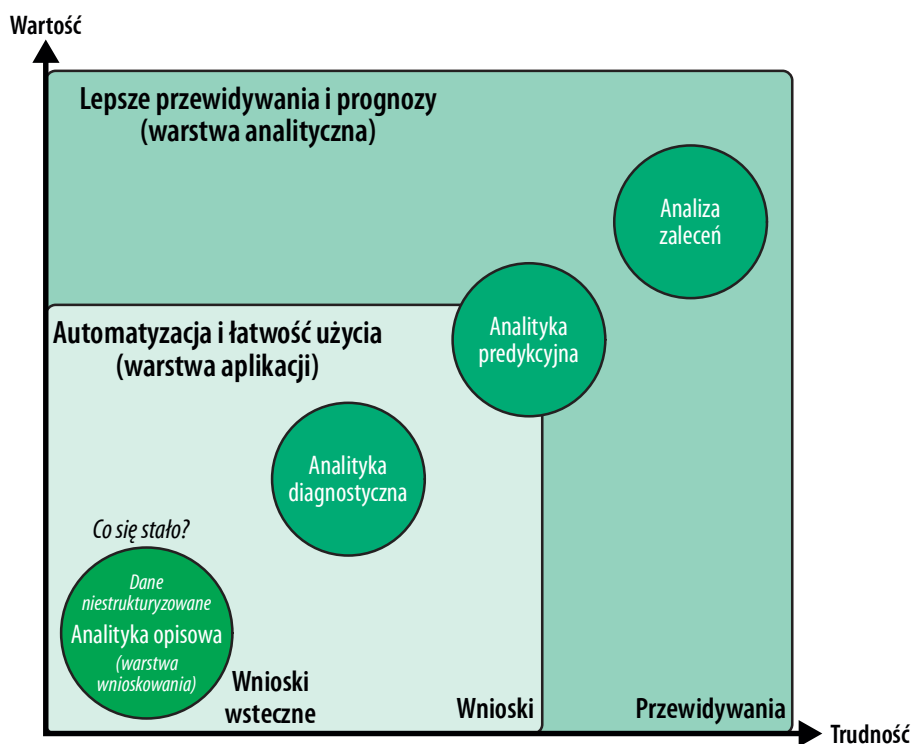
Najbardziej skuteczne zastosowania BI wspomagane przez AI to te, które stapiają ze sobą zautomatyzowane i ludzkie decyzje. Zbadamy to praktycznie w części 2. Teraz przyjrzyjmy się szczegółowo, jak AI może systematycznie pomagać w ulepszaniu naszego rozwiązania BI.

Typowe przypadki użycia AI dla rozwiązań BI

Typowo AI może dodać wartość do rozwiązania BI na trzy sposoby:

- Poprzez zautomatyzowanie wnioskowania i sprawienie, że proces analizowania jest bardziej przyjazny dla użytkownika
- Poprzez obliczanie lepszych prognoz i przewidywań
- Poprzez umożliwienie wyciągania wniosków przez systemy BI nawet z niestrukturalnych źródeł danych

Rysunek 1-3 prezentuje obraz wysokiego poziomu powiązania tych obszarów zastosowań z różnymi metodami analitycznymi.



Rysunek 1-3 Jak możliwości AI wspierają metody analityczne

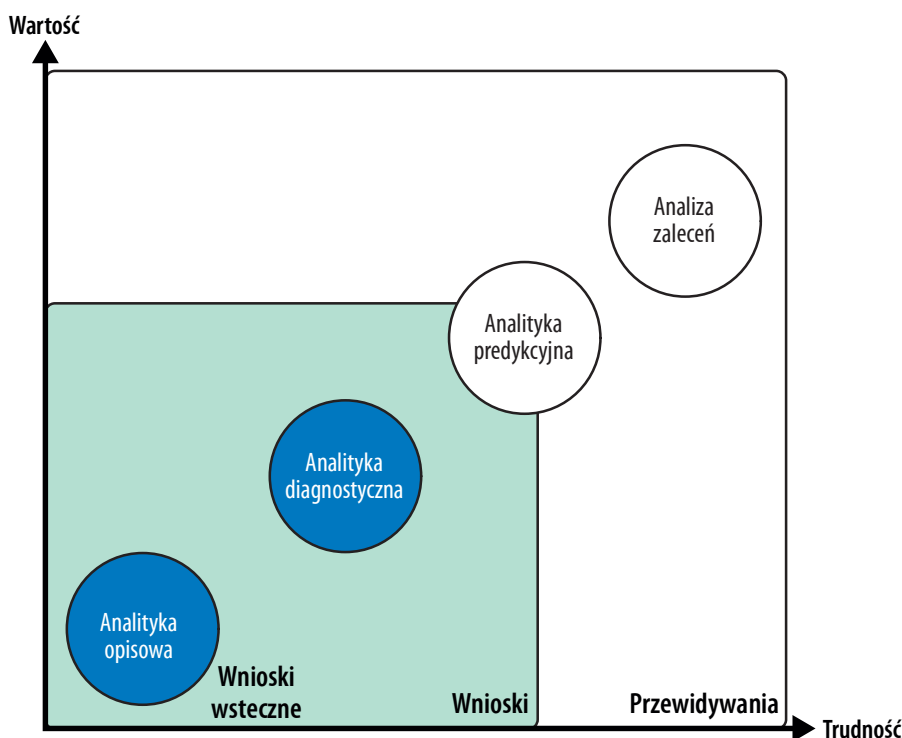
Zbadajmy te obszary nieco dokładniej.

Automatyzacja i łatwość użycia

Uczynienie samego narzędzia BI bardziej inteligentnym i łatwiejszym w użyciu sprawi, że będzie jeszcze bardziej dostępne dla użytkowników nietechnicznych, tym samym zmniejszając obciążenie analityków. Ta łatwość użycia jest zwykle osiągnięta poprzez automatyzację odbywającą się w tle.

Inteligentne algorytmy sprawiają, że możliwe jest szybkie (mierzone w sekundach) przesianie gór danych i wydobywanie interesujących wzorców lub spostrzeżeń dla użytkowników biznesowych lub analityków. Jak widać na rysunku 1-4, procedury te szczególnie dobrze nadają się dla faz analizy opisowej i diagnostycznej.

AI może pomóc w odkrywaniu interesujących korelacji lub nietypowych obserwacji względem wielu zmiennych, które ludzie mogliby przeoczyć. W wielu przypadkach AI jest również lepsza od ludzi w wyszukiwaniu ciekawych kombinacji miar – ludzie zazwyczaj są w stanie skupić się jedynie na jednej metryce na raz. Jednak automatyzacja i łatwość użycia sięga również fazy analizy predykcyjnych – na przykład ułatwiając wytrenowanie i wdrożenie niestandardowych modeli uczenia maszynowego.



Rysunek 1-4 BI napędzane przez AI: automatyzacja i łatwość użycia (warstwa aplikacji)

Trzeba tu zwrócić uwagę na jedną ważną rzecz: funkcjonalności AI w tej fazie są zazwyczaj wbudowane w warstwę aplikacji, którą jest nasze oprogramowanie BI. Zatem typowo nie można po prostu dodać tych możliwości do platformy BI za pomocą paru wierszy kodu Pythona (w przeciwieństwie do zasilanych przez AI prognoz i odblokowywania niestrukturalnych danych, co będziemy omawiali w rozdziałach 7, 8 i 9). Jeśli używamy nowoczesnych platform BI, takich jak Microsoft Power BI albo Tableau, funkcje

wykorzystujące sztuczną inteligencję znajdziemy wewnątrz tych narzędzi. Niekiedy są one ukryte albo odbywają się tak „przezroczyście”, że nawet nie zauważymy, że gdzieś tam pracuje sztuczna inteligencja.

Resztę tego podrozdziału poświęcimy na przedstawienie wskaźników sygnalizujących, że pod maską działa AI, znacznie ułatwiając nasze życie jako analityków.

Wykorzystanie przetwarzania języka naturalnego do interakcji z danymi

Dzięki wykorzystaniu opartej na AI technologii przetwarzania języka naturalnego (*natural language processing* – NLP) komputery znacznie lepiej radzą sobie z interpretowaniem i przetwarzaniem tekstowych poleceń użytkowników. Dla przykładu przypuśćmy, że chcemy poznać wyniki sprzedaży dla ostatniego miesiąca albo wyniki sprzedaży w Stanach Zjednoczonych w minionym roku w porównaniu do bieżącego roku. Moglibyśmy zatem wpisać następujące pytania:

Jakie były sprzedaże w Teksasie w ostatnich 5 latach?

albo

Wartość sprzedaży w Teksasie w minionym roku vs wartość sprzedaży w Teksasie w tym roku

Nie jest potrzebny żaden skomplikowany kod ani specjalizowany język zapytań. Taka warstwa wejściowa w stylu pytań i odpowiedzi (Q&A) sprawia, że platforma BI jest znacznie bardziej dostępna dla użytkowników nietechnicznych, ale również wygodniejsza dla analityków, którzy naprawdę nie są w stanie przewidzieć każdego pytania, które użytkownik biznesowy mógłby chcieć zadać. Większość użytkowników będzie dobrze znać takie podejście, jako że jest podobne do posługiwania się wyszukiwarką, taką jak Google.

Bez względu na to, czy narzędzia Q&A są wbudowane w rozwiązanie BI, czy stanowią odrębną nakładkę, nie wszystkie takie implementacje działają równie dobrze. W istocie niezawodne działanie takich funkcji w środowisku produkcyjnym wymaga rozwiązywania w tle znaczących złożoności. Analitycy muszą śledzić rodzaje pytań zadawanych przez użytkowników biznesowych i upewniać się, że generowane wyjście jest poprawne. Konieczne jest zdefiniowanie synonimów i języka dziedzicznego, aby zagwarantować, że system będzie właściwie interpretować pytania użytkowników. I podobnie jak we wszystkich systemach informatycznych, wymaga to stałego utrzymywania i konserwacji. Mamy jednak nadzieję, że systemy te będą coraz sprawniejsze i ręczne wysiłki potrzebne do ich działania będą malały z czasem.

Podsumowywanie wyników analitycznych

Nawet jeśli treść wykresu wydaje się oczywista i samo-objaśniająca, dobrą praktyką jest podsumowanie kluczowych spostrzeżeń jednym lub dwoma zdaniami języka naturalnego, co zmniejszy ryzyko błędnej interpretacji. Jednak kto naprawdę lubi pisanie pozornie oczywistych objaśnień pod wykresami zamieszczanymi w raportach czy prezentacjach? Większość ludzi tego nie znosi i to jest kolejna sfera, w której AI może pomóc.

NLP napędzane przez AI może nie tylko pomóc w interpretowaniu wejścia w języku naturalnym, ale może również generować tekst podsumowania na podstawie przetwarzanych danych. Takie automatycznie generowane teksty będą zawierać opisowe charakterystyki danych, a także godne uwagi zmiany lub trendy. Oto przykład podpisu wykresu automatycznie wygenerowanego przez Power BI:

Wartość Sales w Texas wzrastała w ostatnich 5 latach i wykazuje najdłuższy okres wzrostowy dla Sales pomiędzy 2010 a 2014

Jak można zauważyć, takie drobne, generowane przez AI fragmenty tekstu mogą znacząco ułatwić pracę analityka i oszczędzić mnóstwo czasu, gdy przychodzi czas na zakomunikowanie spostrzeżeń innym zainteresowanym. Dodatkowo mogą one pomóc w spełnianiu wymogów dostępności, tworząc tekst dla czytników ekranowych.

Wykorzystanie automatyzacji do wyszukiwania wzorców w danych

Zobaczyliśmy już, że funkcjonalność NLP może pomóc w wydajnym uzyskaniu opisowych spostrzeżeń z naszych danych. Kolejnym logicznym krokiem jest odkrycie, dlaczego konkretne zjawisko wystąpiło w przeszłości, na przykład dlaczego sprzedaż w Teksasie wzrosła tak bardzo?

W przypadku analiz diagnostycznych normalnie musielibyśmy przeczesać nasz zbiór danych, aby zbadać znaczące zmiany w rozkładach danych. W tym przykładzie możemy zechcieć ustalić, czy ogólne zmiany były napędzane określonym produktem lub zdarzeniem. Ten proces szybko może stać się żmudny i kłopotliwy. AI może pomóc w zmniejszeniu *czasu wnioskowania* (*time to insight* – TTI).

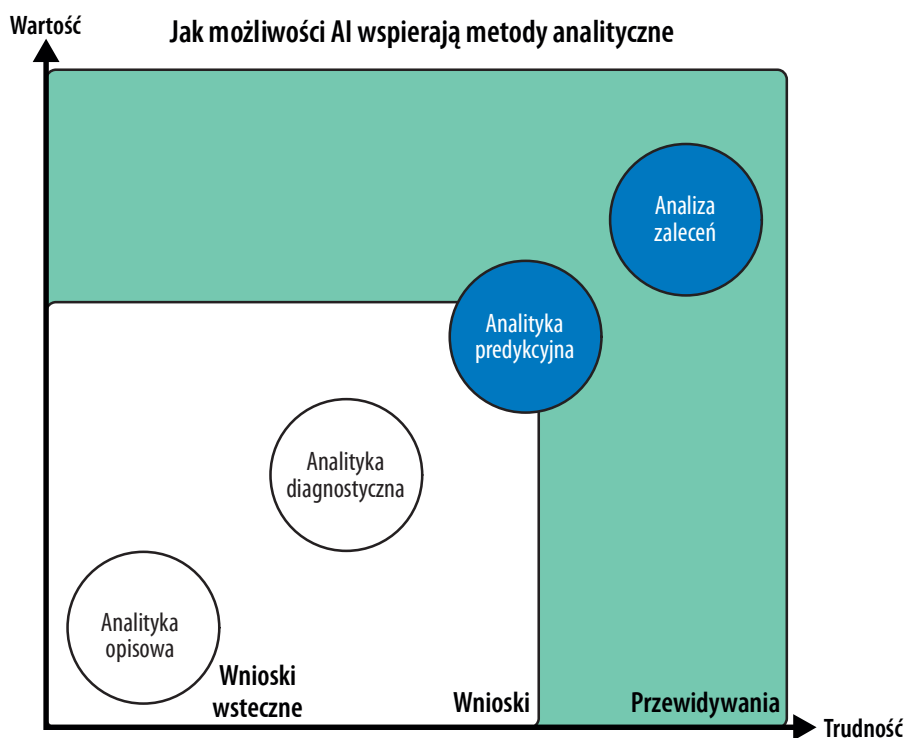
Algorytmy doskonale radzą sobie z wykrywaniem wzorców w danych w tle i wydobywaniem ich na pierwszy plan. Dla przykładu, za pomocą narzędzi sztucznej inteligencji, takich jak drzewa rozkładu lub analiza kluczowych czynników, można szybko ustalić, jakie charakterystyki w naszych danych doprowadziły do ogólnego zaobserwowanego efektu – w locie. W rozdziałach 5 i 6 przyjrzymy się trzem konkretnym przykładom wykorzystania wspieranych przez AI możliwości w Power BI, dzięki którym nasze życie jako analityka danych lub użytkownika biznesowego będzie lżejsze.

Lepsze prognozowanie i przewidywania

Podczas gdy analizy opisowe i diagnostyczne stanowią serce każdego systemu BI, od zawsze istniało też pragnienie, aby nie tylko zrozumieć przeszłość, ale również przewidywać przyszłość. Jak można zobaczyć na rysunku 1-5, możliwości wzmacniane sztuczną inteligencją mogą wspomagać końcowych użytkowników w stosowaniu skutecznych metod prognostycznych i analityki zaleceń w celu lepszego *prognozowania* i lepszych *przewidywań* opartych na danych historycznych.

Powoduje to wzrost złożoności, gdyż opuszczamy królestwo binarnych danych z przeszłości i wprowadzamy probabilistyczne domysły na temat przyszłości, które oczywiście zawierają wiele niepewności. Jednocześnie perspektywiczna wartość rośnie: jeśli

jesteśmy w stanie przewidzieć przyszłość, możemy podejmować znacznie lepsze decyzje w teraźniejszości.



Rysunek 1-5 BI napędzane przez AI: lepsze prognozy i przewidywania (warstwa analityczna)

Być może czytelnik słyszał wcześniej o metodach statystycznych, takich jak regresja lub autoregresyjna zintegrowana średnia krocząca (*autoregressive integrated moving average* – ARIMA) – być może w szkole średniej lub na podstawowych wykładach uniwersyteckich – i zastanawia się, gdzie tu jest miejsce dla AI. Trzeba jednak zwrócić uwagę na następujące aspekty:

AI może tworzyć lepsze prognozy, wykorzystując więcej danych i mniejszy nadzór ze strony człowieka

AI w swoich podstawach wykorzystuje dobrze znane, sprawdzone techniki, takie jak regresja liniowa. Jednocześnie jednak AI może stosować te techniki do złożonych zbiorów danych, wykorzystując podejścia stochastyczne do szybkiego odszukania optymalnego rozwiązania bez konieczności obszernego nadzoru ludzkiego. Specjalizowane algorytmy dla prognoz serii czasowych są zaprojektowane w celu rozpoznawania wzorców w większych ilościach takich danych. AI próbuje optymalizować przewidywania bazując na wyborze cech i minimalizowaniu funkcji straty. Może to doprowadzić do lepszych lub dokładniejszych przewidywań dla krótkiego horyzontu czasowego albo próbowania uzyskania bardziej dokładnych prognoz dla dłuższych okresów czasu. Bardziej złożone, nieliniowe modele pozwalają uzyskać bardziej szczegółowe, a w konsekwencji lepsze wyniki predykcji.

AI może obliczać prognozy na dużą skalę w celu zoptymalizowania podejmowania decyzji

Przewidzenie całkowitej liczby klientów w następnym kwartale byłoby miłe. Co jednak byłoby jeszcze lepsze, to obliczenie prawdopodobieństwa rotacji dla każdego klienta z naszej bazy, bazując na ostatnich danych. Mając te informacje, możemy nie tylko ustalić, którzy klienci najprawdopodobniej odejdą w następnym miesiącu, ale również zoptymalizować proces podejmowania decyzji. Możemy dla przykładu ustalić, którzy spośród klientów zapewne zechcą odejść w kolejnym miesiącu i do nich właśnie skierować kampanię marketingową. Połączenie uczenia maszynowego z BI tworzy potencjalnie ogromne możliwości dla organizacji. Zaś wraz z rozwojem najnowszych technik, takich jak zautomatyzowane uczenie maszynowe (AutoML) oraz sztuczna inteligencja jako usługa (*AI as a service* – AIaaS), które zbadamy bliżej w rozdziale 3, organizacje mogą zredukować wąskie gardła powodowane przez brak dostatecznej liczby specjalistów od danych lub praktyków ML, utrudniające wykorzystanie potencjału AI.

Możliwości AI ulepszające prognozowania lub lepsze przewidywania można napotkać jako integralną część istniejącego oprogramowania BI (warstwa aplikacji). Możliwości te mogą być również stosowane niezależnie, bezpośrednio na poziomie bazy danych (warstwa analityczna). Sprawia to, że zawsze są dostępne, bez względu na to, jakiego narzędzia BI używamy. Stosowanie tych technik zbadamy bliżej w rozdziałach 7 i 8.

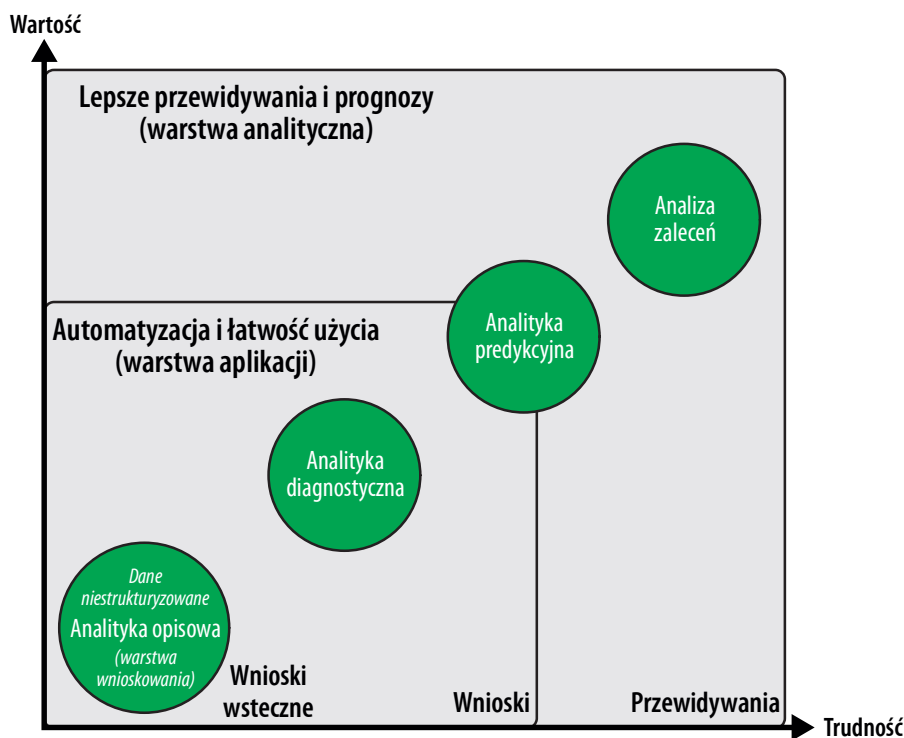
Posługiwanie się danymi niestrukturyzowanymi

Systemy BI typowo pracują z danymi tabelarycznymi z relacyjnych baz danych, takich jak hurtownie danych przedsiębiorstwa. Ostatnio jednak, wraz z rosnącą cyfryzacją wszelkich kanałów, można zaobserwować ogromny wzrost użycia danych niestrukturalnych, mających formę plików tekstowych, obrazów lub audio. Historycznie rzecz ujmując, te formaty danych są trudne do analizowania w większej skali. Sztuczna inteligencja pozwala to zmienić.

AI może poszerzyć i pogłębić dane dostępne dla odczytu przez maszyny, wykorzystując takie technologie, jak widzenie komputerowe lub NLP, aby uzyskiwać dostęp do nowych, wcześniej nieuwzględnianych źródeł danych. *Niestrukturalne dane*, takie jak pliki tekstowe, dokumenty PDF, obrazy czy pliki audio można przekształcić w formaty strukturyzowane pasujące do zadanego schematu, takiego jak tabela lub plik CSV, a następnie mogą być one używane i analizowane w systemie BI. Jako że jest to coś, co następuje na poziomie pozyskiwania danych, proces ten nieuniknienie wpływa na wszystkie fazy działania platformy BI (rysunek 1-6).

Dzięki włączeniu tych plików do naszych analiz możemy uzyskać jeszcze więcej informacji, które potencjalnie mogą doprowadzić do lepszych prognoz lub lepszego rozpoznania kluczowych czynników.

W rozdziale 8 zobaczymy na przykładach, jak to działa w praktyce.



Rysunek 1-6 BI napędzane przez AI: odblokowanie niestrukturalnych danych w warstwie pozyskiwania danych

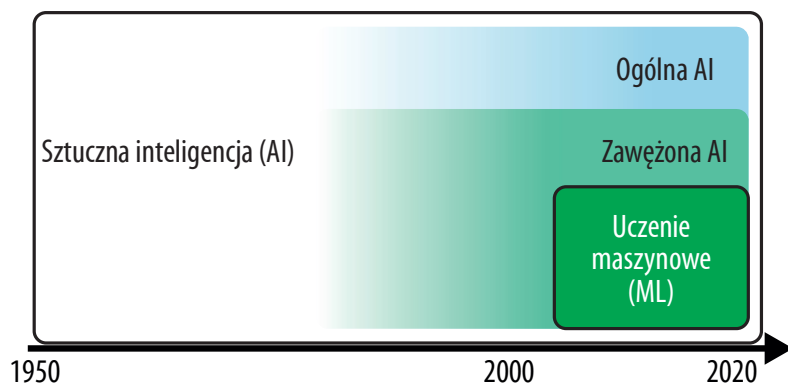
Uzyskanie intuicyjnego zrozumienia AI i uczenia maszynowego

Powiedzieliśmy już sporo na temat tego, jak można wykorzystać sztuczną inteligencję w rozwiązaniach BI. Jednak by móc naprawdę budować produkty lub usługi wspomagane przez AI, musimy pogłębić rozważania i zrozumieć, czym AI jest i co pozwala (a także czego nie pozwala) osiągnąć.

Zatem czym naprawdę jest AI? Jeśli zapytamy o to dziesięć osób, zapewne otrzymamy jedenaście odpowiedzi. Na potrzeby wykładu prezentowanego w tej książce ważne jest uzgodnienie pojęć i ustalenie, co ten term rzeczywiście oznacza.

Na początek, trzeba zauważyć, że termin *sztuczna inteligencja* nie jest nowy. W istocie był używany w wojskowych laboratoriach badawczych już w latach pięćdziesiątych XX wieku. Od tego czasu badacze podejmowali wiele prób osiągnięcia celu, którym było zreplikowanie ludzkiej inteligencji przez maszyny (komputery). Jak pokazuje rysunek 1-7, od czasu wymyślenia tego pojęcia rozwinęły się dwa ogólne pola sztucznej inteligencji: ogólna AI oraz zawężona AI.

Ogólna AI, nazywana też *silną AI*, dotyczy technologii, która dąży do rozwiązania dowolnego zadanego problemu, z którym system nigdy wcześniej nie miał styczności, podobnie do tego, jak działa ludzki umysł. Ogólna AI pozostaje gorącym tematem badawczym, ale cel ciągle wydaje się odległy; badacze nadal nie mają pewności, czy w ogóle kiedykolwiek zostanie osiągnięty.



Rysunek 1-7 *Rozwój sztucznej inteligencji*

Z drugiej strony *zawężona AI*, nazywana też *słabą AI*, dotyczy raczej konkretnego rozwiązania, które jest w stanie radzić sobie z pojedynczym, dobrze zdefiniowanym problemem, dla którego zostało zaprojektowane i wytrenowane. Słaba AI napędza wszystkie przełomowe osiągnięcia AI, które obserwujemy w ostatnich latach, zarówno w obszarach badawczych, jak i zastosowaniach praktycznych (biznesowych).

U podstaw *zawężonej AI* wyróżnia się jedno podejście, jeśli chodzi o wpływ biznesowy i postępy projektowe: *uczenie maszynowe (machine learning – ML)*¹. W istocie, ilekroć w tej książce wspominam o AI, mamy do czynienia z rozwiązaniem, które stało się możliwe dzięki ML. To dlatego będę zamiennie używać terminów *AI* oraz *ML* w tej książce. Można uznać, że AI jest raczej szerszym pojęciem o całkiem dosłownym znaczeniu: AI jest narzędziem do budowania (pozornie) inteligentnych bytów, które są w stanie rozwiązywać konkretne zadania, głównie za pomocą ML.

Teraz, gdy zależność pomiędzy AI i ML jest nieco jaśniejsza (mam nadzieję), zajmijmy się tym, czego naprawdę dotyczy uczenie maszynowe. *ML to paradygmat programistyczny nakierowany na wyszukiwanie w danych wzorców, na potrzeby określonego zastosowania*. ML typowo obejmuje dwie fazy: *uczenie się* (trening) oraz *wnioskowanie* (nazywane też *testowaniem* lub *przewidywaniem*).

Kluczowa koncepcja leżąca u podstaw ML polega na tym, że w historycznych danych wyszukujemy wzorce, aby rozwiązać określone zadanie, takie jak rozdzielenie obserwacji na różne kategorie, oszacowanie prawdopodobieństw albo wyszukanie podobieństw pomiędzy różnymi bytami. Typowy przypadek użycia ML to analizowanie historycznych danych transakcyjnych w celu obliczenia indywidualnych prawdopodobieństw rotacji klientów. W fazie wnioskowania naszym celem jest obliczenie prognozowanej wartości nowego punktu danych na podstawie tego wszystkiego, czego nauczyliśmy się z danych historycznych.

Aby pogłębić zrozumienie ML, przyjrzymy się kluczowym składowym naszej definicji:

1 Polski termin jest dość niezręczny, gdyż można go zrozumieć jako „uczenie za pomocą maszyn”, podczas gdy w istocie chodzi o „uczenie SIĘ maszyn”. Jednak zwrot „uczenie maszynowe” rozpowszechnił się już i trzeba się z tym pogodzić (przyp. tłum.).